

令和 5 年 6 月 4 日現在

機関番号：17102

研究種目：若手研究

研究期間：2020～2022

課題番号：20K19771

研究課題名（和文）単純再帰型ニューラルネットワーク向け光コンピューティングシステム基盤の研究

研究課題名（英文）Nano-photonic Processing Unit for Recurrent Neural Network Applications

研究代表者

川上 哲志（Kawakami, Satoshi）

九州大学・システム情報科学研究所・准教授

研究者番号：20845523

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：本研究は、記憶機能を有する単純再帰型ニューラルネットワーク向け光演算回路を基本とした高性能・低電力な光計算機システムを確立することを目的とし光電融合演算器システムを提案した。提案回路は、再帰経路の位相差を校正する機構と光電融合非線形活性化関数を導入することで、RNNアプリケーションを効率的に実行する演算器を実現した。提案回路の精度調査においては、ノイズによる演算誤差が発生する環境下においても実アプリが精度劣化なく推論できることを確認した。さらに、既存の光/電気アクセラレータと比較し、それぞれ467倍・7.3倍の性能改善、93%・58%の省エネルギー化が達成できることを確認した。

研究成果の学術的意義や社会的意義

本研究課題の学術的意義は、ポストムーア時代のコンピュータ構成法として光デバイス活用のポテンシャルを示すことにある。現在の光演算器研究は、信号光波が回路内を1方向に1回だけ通過する演算方式を基本としており、その最大の特徴は光速性を活かした低遅延性である。しかしながら、低遅延な演算回路はより多くのメモリアクセスとOE/AD変換を要求し、システム性能を悪化させる可能性がある。本研究では、信号光波が複数回通過可能な再帰経路を有する演算器を基本とすることで、性能的にも消費エネルギー的にも優位であることを示した。これは、光デバイスを活用した次世代計算機の実現に向けた基盤技術創出へつながるものである。

研究成果の概要（英文）：This work proposes an optoelectronic processing system based on an optical arithmetic circuit for simple recurrent neural networks with memory function, aiming to establish a high-performance and low-power optical computer system. The proposed circuit has a mechanism to calibrate the phase difference of the recursive path and an OEO nonlinear activation function to realize an operator that can efficiently execute RNN applications. In the accuracy evaluation, we confirmed that the real application can be inferred without degradation of accuracy even in an environment where arithmetic errors are due to noise. Furthermore, we confirmed that the proposed circuit could achieve 467x and 7.3x performance improvement and 93% and 58% energy saving compared to existing optical/electrical accelerators, respectively.

研究分野：計算機システムアーキテクチャ

キーワード：光コンピューティング 再帰型ニューラルネットワーク

1. 研究開始当初の背景

情報処理性能向上の中核を担う半導体集積回路の微細化の終焉に伴い、従来の CMOS デジタル回路によるノイマン型コンピュータに替わる新たな計算機技術が世界で求められている。その中、ナノフォトニックデバイスを活用した光コンピューティングは、計算機の継続的高性能化・低消費電力化を支える有望な代替コンピューティングアプローチの1つである。近年の光デバイスの高密度集積化技術(ナノフォトニクス)の発展により、光デバイスの超小型化、超低消費エネルギー化、大規模集積化が可能になり、情報通信技術への適応が主だった光技術を情報処理へ応用する機運となった。事実、光加算器や学術/産業界ともに注目を浴びているフィードフォワードニューラルネットワーク (FFNN) 処理向けの専用光回路が提案されており、基礎回路の動作検証により実現可能性が示されている。

一方で、再帰型ニューラルネットワーク処理向けの光回路の検討は少なく、特に「光学特性に起因する大容量光メモリの実現の困難性」や「現代社会を支える様々な汎用計算機が電気回路で構成される」ことを鑑みると、電気メモリの活用や光電(OE)変換/アナログ デジタル(AD)変換は必須である。如何に優れた(低遅延な)光デバイスや演算回路を実現したとしてもメモリ通信や OE/AD 変換がボトルネックとなれば、計算機システムとしての恩恵(高性能/低電力性)を享受できない。したがって、これまでの光デバイス/回路分野に閉じた議論から脱却し、電気メモリやインターフェースを含めたシステム全体を俯瞰した上で光回路のあるべき姿を再検討することで、より光の特性を活かした計算機基盤を確立することが重要である。

2. 研究の目的

本研究の目的は、「ポストムーア時代のコンピュータ構成法として光デバイス活用のポテンシャルを示す」ことにある。現在の光演算器研究は、信号光波が回路内を1方向に1回だけ通過する演算方式を基本としており、その最大の特徴は光速性を活かした低遅延性である。しかしながら、低遅延な演算回路はより多くのメモリアクセスと OE/AD 変換を要求し、システム性能を悪化させる可能性がある(図1)。したがって、システムレベルでの利用を見据えた光演算回路の再検討が必要である。本研究では、信号光波が複数回通過可能な再帰経路を有する演算器を基本とすることで、光速性は活かしつつ当該オーバーヘッドの削減を狙う。これにより、システムレベルで高性能/低電力な光計算機基盤を構築することを目的とする。

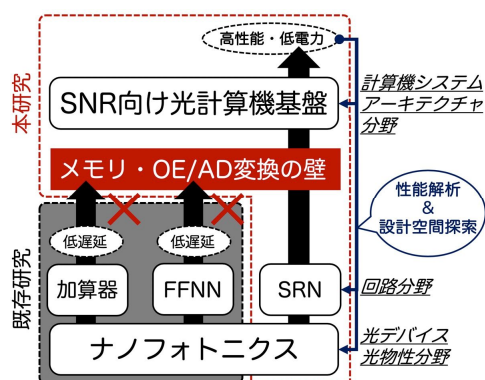


図1. 本研究の概要

3. 研究の方法

本研究では単純再帰型ニューラルネットワーク (SRNN) 向けの光計算機を前提とし、その光演算回路の確立を目指す。SRNN は、FFNN と異なり再帰層を有するため再帰経路型の演算回路と親和性が高い。また、動的エネルギーロスが小さいという光特有の性質も再帰回路に適する。さらに、再帰層は短時間の記憶機能を有するため、回路構成や処理分配法次第でメモリアクセスを削減できる可能性がある。したがって、SRN を対象とすることは、単なる応用アプリケーション拡大のみならず、メモリを含めた光計算機システムの性能/電力の改善機会を提供する。

上記を達成すべく、下記3つの実施項目を設定する。

(1) SRNN 向け光演算回路の構築

はじめに SRNN 向けの光演算回路を構築する。SRNN も FFNN と同様に主要な演算は行列積であるためその知見を基礎として、SRN 特有の再帰層の演算回路を拡張する。具体的には、光行列ベクトル積演算回路である MZI-VVM をベースとして再帰経路を有する SRNN 向けの光電融合演算回路へと拡張する。本研究では、再帰経路の構築のための位相調整機構と光演算器による活性化関数の実現に取り組む。

(2) 演算精度とアプリケーション推論精度解析

本提案光演算回路は、アナログ演算を基本としているため演算精度劣化が懸念される。加えて、アナログ演算を繰り返す事による誤差の累積とそれによる精度の悪化も考えられる。したがって、当該回路の実現可能性を検証するためには、精度検証が必須となる。本研究では、まず演算器としての計算の誤差を示す「演算精度」を評価した後、実アプリケーションによる「推論精度」を評価することによって、演算器のポテンシャルのみならず実用上の誤差も評価する。

(3) システム性能(レイテンシー)・消費エネルギーの解析

光デバイスは本質的に CMOS に対する優位性を有するが、本デバイスによって構成された演算器がシステムレベルで優位性をどの程度持つか不明である。したがって、システムレベルでのレイテンシー・消費エネルギーの評価を実施することで、提案回路の優位性を定量化する。

4. 研究成果

(1) SRNN 向け光演算回路の構築

当該ハードウェアは、可能な限り全ての演算を光回路で完結させることで、高性能（低レイテンシ）化を指向するものである。光再帰経路の導入に際し、光を情報媒体として循環させるためニューラルネットワークにおける全ての演算を光回路で実現する必要がある。したがって、「従来の光演算回路には無い光循環経路」と「対象とする演算を光回路内で完結する必要があるための光活性化関数の構築」が必須となる。提案する光回路を図2に示す。本回路は、単純再帰型ニューラルネットワーク（SRNN）を実行するための光電融合回路である。したがって、処理対象となるSRNNの処理を陽に実行する回路が必要となる。例えば、SRNNを構成するニューロン数が、入力層-再帰層-出力層でそれぞれ2-4-1である場合を考えると、 2×4 （入力層-再帰層間シナプス）/ 4×4 （再帰層-再帰層間シナプス）/ 4×1 （再帰層-出力層間シナプス）の行列-ベクトル積演算回路と各中間結果に対する非線形演算回路が必要となる。行列-ベクトル積演算を実行するための光回路実装は各種研究提案があり、本研究ではマツハツェンダ干渉器をアレイ状に構成したMZI-VMMを活用する。

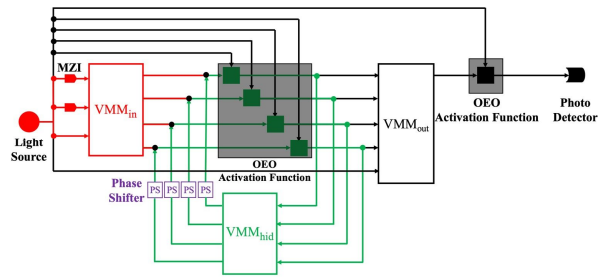


図2. 再帰型光演算回路

本研究回路の重要な点の1つが再帰経路に挿入されている位相シフターである（図2紫）。光循環経路を実装する場合光の干渉により加算演算が実行されることを期待するが、とある時刻に入力された光波（図2赤）と前時刻に実行された演算の中間結果を示す光波（図2緑）の位相は通常異なるため干渉により加算演算が実行されることは稀である。例えば、2つの信号光波の位相が相対的に π ズレている場合、光の干渉により減算演算が実行される。この問題を解消すべく本提案回路において位相シフターを導入した。これによりいかなる循環経路を有する光回路においても位相差を解消し所望の加算演算を実行可能である。さらに、デバイスの製造ばらつきに起因するアナログ演算誤差も同時にこの位相シフターで較正可能である。

次に、対象とする演算を光回路内で完結する必要があるための光活性化関数を構築した。非線形応答を持つ光デバイスは、過飽和吸収体や2光子吸収素子など各種存在するが、その入出力応答や伝達効率の観点から本演算回路には向いていない。そこで、ニューラルネットワーク向け活性化関数であるReLU関数の実現を目指した。当初、光回路シミュレータ上で光ReLU関数の動作を確認したが、当該回路には光乗算回路を必要としており実装困難性の観点から、光活性化関数の変更を検討した。変更後の光活性化関数ではReLU関数の特徴である、負の数の切り捨て機能を実現するためのコヒーレントレシーバーと再帰経路による光の減衰を防ぐための強度増幅機能を備えたOEO（光電光）変換器から構成される（図3）。OEO-AFに入射された光波は、コヒーレントレシーバーを介して電流値に変換される。この際、基準となる参照光に対して逆位相（ π ）の光波はマイナスの電流値として検出される。変換された電流値はポンプレーザーの制御信号として扱われるため、負の電流値は切り捨てられ、正の電流値に比例した強度の光波として出力される。本回路は光波の電解振幅に対し情報がエンコードされるため、OEO-AFの入出力電解振幅応答はルート関数となる。これは、当初目標としていたReLU関数の応答と異なるため計算機サーバー上で学習可能性を検証し、実アプリケーションにおいて学習可能であることを確認した。以上により、SRNN向けの光電融合演算回路が構築できた。

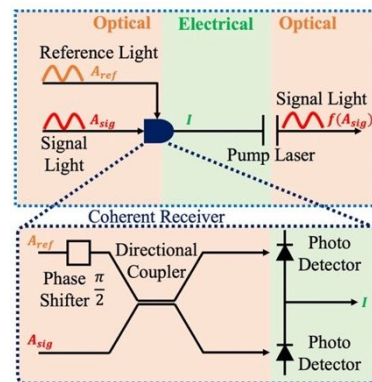


図3. OEO 活性化関数

変換された電流値はポンプレーザーの制御信号として扱われるため、負の電流値は切り捨てられ、正の電流値に比例した強度の光波として出力される。本回路は光波の電解振幅に対し情報がエンコードされるため、OEO-AFの入出力電解振幅応答はルート関数となる。これは、当初目標としていたReLU関数の応答と異なるため計算機サーバー上で学習可能性を検証し、実アプリケーションにおいて学習可能であることを確認した。以上により、SRNN向けの光電融合演算回路が構築できた。

(2) 演算精度とアプリケーション推論精度解析

(1)で構築した光電融合演算回路は、アナログ演算をベースとしており演算精度の劣化が考えられる。この影響を正確に解析すべく、「演算器としての計算精度を評価するための「演算精度評価」と「SRNN実アプリを実行した際の推論精度評価」を実施した。精度評価実験は、光回路シミュレータ(Optisystem)上に実装した提案回路の入力光源にガウシアンノイズをインジェクションすることで行う。

の演算精度評価に於いては、入出力器は8bitの量子化（定義域：0~255）が行われることとし、量子化されたデジタル値を真値と比較することによって誤差を測定する。誤差を最大限観測すべく出力理論値が定義域の中央である128となるような入力値・行列値を固定して演算を実行する。演算は一定のサンプリング周期ごとにシリアルに入力されるものとし、再帰回数20回まで計算を続ける。また、ノイズ源となるガウス分布の分散を 10^{-6} 、 10^{-9} 、 10^{-12} 、 10^{-15} と変化させて実験を実施した。図4にOptisystemから得られる光電融合回路の計算値と計算機サーバー

上で実行された観測値の誤差の平均値と標準偏差を示す。結果をから、計算結果の平均値は0付近を遷移していることわかる。つまり、ReLUのような負切り捨て活性化関数を実行した場合においてもアナログ誤差は0に集約されることがわかった。一方で、誤差の標準偏差は再帰回数（横軸）が増えるにつれ増加していることがわかる。これは、本提案回路において誤差が累積し計算が破綻することを示唆する結果となった。

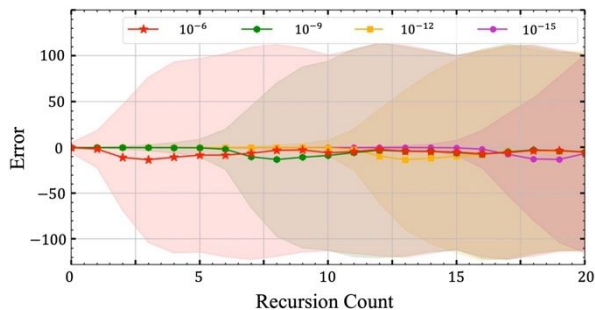


図 4. 演算精度評価における誤差

のアプリケーション推論精度評価においては、16 bit serial adder を実装した。本アプリは、正確な演算のためには前時刻の情報（キャリー情報）を記憶する必要があり SRNN の機能を検証するものの1つとして適切である。アプリは計算機サーバー上で、Rmsprop アルゴリズム上で学習し、デジタル計算機上で推論精度 100% となったことを確認した後、Optisystem 上で光電融合回路実行を検証した。その結果、提案回路でのノイズを考慮した実行においても推論精度 100% となることが判明した。図 4. の結果においては再帰回数が 16 回に達した際に 10^{-6} , 10^{-9} , 10^{-12} のノイズ源において計算が破綻することが予想されるが、その結果は得られなかった。詳細な解析のために、アプリ実行の際の誤差評価を実施した。図 5 は、再帰回数の増加にしたがって演算誤差が累積していないことを示す。この結果は、図 4 と異なりアプリ側の学習データが過去の信号を忘却するような推論を進めている事がわかる。つまり、アプリケーション側がある程度の誤差を許容（吸収）できる可能性を示すものである。以上により、提案回路は一定の誤差は発生するものの実アプリでの推論も十分可能であることが示された。

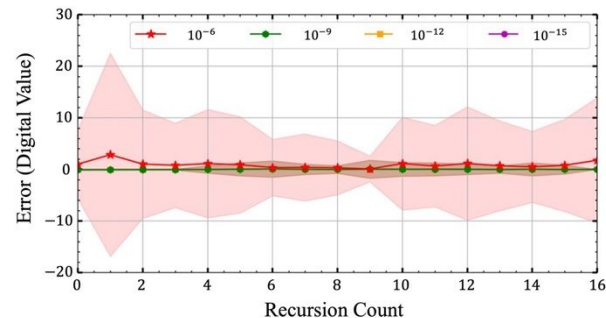


図 5. 推論精度評価における誤差

(3) システム性能（レイテンシー）・消費エネルギーの解析
 本提案回路の優位性を定量的に示すべくレイテンシー評価と消費エネルギー評価を実施した。比較対象となる演算回路はを図 6 に示す。(a)は提案回路、(b)はアナログ光演算回路は活用するが再帰経路を有さない既存のフィードフォワード型光アクセラレータであり、(c)は CMOS デジタルのニューラルネットワーク向けのストリクアレイアクセラレータ(Google TPU など)を示す。

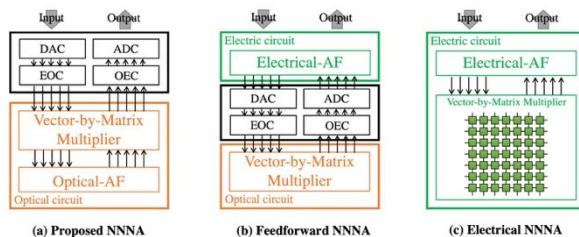


図 6. 比較対象計算機回路一覧

図 7 に各アクセラレータのレイテンシー比較と内訳を示す。(b)既存の光アクセラレータに対し約 7 倍、(c)CMOS アクセラレータには 467 倍の性能向上を達成できることがわかった。特に、(b)で多くの時間を要している OE/E0 変換、DA/AD 変換のオーバーヘッドを削減できている事がわかり、本研究提案の再帰経路の恩恵が大きいことが確認できた。

図 8 に各アクセラレータの消費エネルギー比較と内訳を示す。提案回路は、(b)に対し約 2.4 倍、(c)に対し約 16 倍の効率性を持つことがわかる。レイテンシー評価と同様に(b)との内訳を比較すると OE/E0 変換、DA/AD 変換のオーバーヘッドを削減により提案回路が優位性を示していることがわかる。以上により、提案回路は性能・消費エネルギーの両面で既存研究に対し大きな優位性を示すことが判明した。

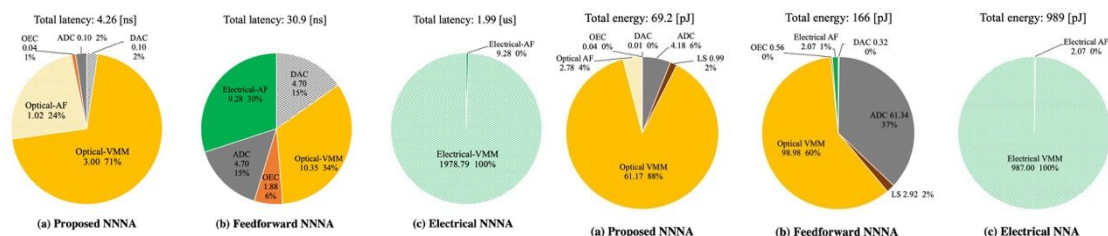


図 7. レイテンシー評価

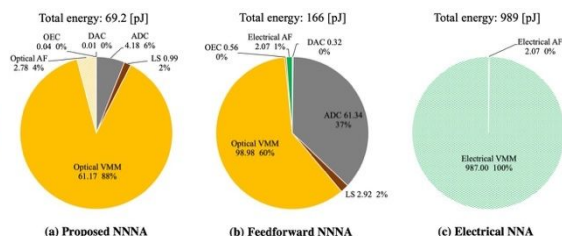


図 8. 消費エネルギー評価

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計24件（うち招待講演 13件 / うち国際学会 7件）

1. 発表者名 Eito Sato, Koji Inoue and Satoshi Kawakami
2. 発表標題 Design and Analysis of a Nano-photonic Processing Unit for Low-Latency Recurrent Neural Network Applications
3. 学会等名 IEEE International Symposium on Embedded Multicore/Many-core Systems-on-Chip (国際学会)
4. 発表年 2022年

1. 発表者名 Takumi Inaba, Takatsugu Ono, Koji Inoue and Satoshi Kawakami
2. 発表標題 A Hybrid Opto-Electrical Floating-point Multiplier
3. 学会等名 IEEE International Symposium on Embedded Multicore/Many-core Systems-on-Chip (国際学会)
4. 発表年 2022年

1. 発表者名 Satoshi Matsushita, Teruo Tanimoto, Satoshi Kawakami, Takatsugu Ono, Koji Inoue
2. 発表標題 An Edge Autonomous Lamp Control with Camera Feedback
3. 学会等名 IEEE World Forum on Internet of Things (WF-IoT) (国際学会)
4. 発表年 2022年

1. 発表者名 Kuan Yi Ng, Aalaa M.A. Babai, Satoshi Kawakami, Teruo Tanimoto and Koji Inoue
2. 発表標題 Layer-wise power/performance analysis for single-board CNN inference
3. 学会等名 The cross-disciplinary Workshop on Computing Systems, Infrastructures, and Programming
4. 発表年 2022年

1. 発表者名 Iori Ishikawa, Koki Ishida, Masamitsu Tanaka, Ikki Nagaoka, Satoshi Kawakami, Teruo Tanimoto, Akira Fujimaki and Koji Inoue
2. 発表標題 Design of Variable Bit-Width Arithmetic Unit Using Single Flux Quantum Device
3. 学会等名 IEEE International Symposium on Circuits & Systems (国際学会)
4. 発表年 2022年

1. 発表者名 川上 哲志
2. 発表標題 フォトリックコンピューティングのための光系・電気系タスク分解論
3. 学会等名 電子情報通信学会総合大会 ソサイエティ特別企画セッション (招待講演)
4. 発表年 2023年

1. 発表者名 北山 研一, 山本 裕紹, 天野 英晴, 納富 雅也, 川西 哲也, 川上 哲志, 笠松 章史, 成瀬 誠
2. 発表標題 フォトリックコンピューティングへのアプローチ - 光の極限性能に迫る -
3. 学会等名 光の極限性能を生かすフォトリックコンピューティングの創成第1回公開シンポジウム (招待講演)
4. 発表年 2023年

1. 発表者名 川上 哲志
2. 発表標題 光コンピューティングシステムアーキテクチャの課題と可能性
3. 学会等名 Optics & Photonics Japan (招待講演)
4. 発表年 2022年

1. 発表者名 Masamitsu Tanaka, Ikki Nagaoka, Satoshi Kawakami, Teruo Tanimoto, Takatugu Ono, Koji Inoue, and Akira Fujimaki
2. 発表標題 High-Throughput Single-Flux-Quantum Circuits Based on Gate- Level-Pipelining toward Artificial Intelligence Applications
3. 学会等名 The Superconducting SFQ VLSI Workshop (招待講演) (国際学会)
4. 発表年 2022年

1. 発表者名 Masamitsu Tanaka, Ikki Nagaoka, Satoshi Kawakami, Teruo Tanimoto, Takatugu Ono, Koji Inoue, and Akira Fujimaki
2. 発表標題 Demonstration of a Superconductor 8-Bit Microprocessor Based on High-Throughput Single-Flux-Quantum Logic Circuits
3. 学会等名 The East Asia Symposium on Superconductor Electronics (招待講演) (国際学会)
4. 発表年 2023年

1. 発表者名 Aalaa M.A. Babai, Kuan Yi Ng, Teruo Tanimoto, Satoshi Kawakami and Koji Inoue
2. 発表標題 Non-Volatile FPGA-based Intermittent Computing and Its Performance Analysis
3. 学会等名 情報処理学会研究報告
4. 発表年 2022年

1. 発表者名 佐藤 英人, 川上 哲志, 岡 慶太郎, 谷本 輝夫, 小野 貴継, 井上 弘士
2. 発表標題 単純再帰型ニューラルネットワーク向け光演算回路の初期検討
3. 学会等名 LSIとシステムのワークショップ
4. 発表年 2021年

1. 発表者名 上野 麟, 谷本 輝夫, 後藤 孝行, 丸岡 晃, 川上 哲志, 小野 貴継, 飯塚 拓郎, 井上 弘士
2. 発表標題 Halideを用いたオイラー動画像誇張処理のCPU-FPGAハイブリッドシステムの設計と実装
3. 学会等名 LSIとシステムのワークショップ
4. 発表年 2021年

1. 発表者名 上野 麟, 谷本 輝夫, 後藤 孝行, 丸岡 晃, 川上 哲志, 小野 貴継, 飯塚 拓郎, 井上 弘士
2. 発表標題 オイラー動画像誇張処理を対象とした CPU-FPGAハイブリッドシステムの実装と評価
3. 学会等名 組込み技術とネットワークに関するワークショップ
4. 発表年 2021年

1. 発表者名 佐藤英人, 川上 哲志, 岡 慶太郎, 谷本 輝夫, 小野 貴継, 井上 弘士
2. 発表標題 単純再帰型ニューラルネットワーク向けナノフォトニックアクセラレータの設計
3. 学会等名 並列 / 分散 / 協調処理に関するサマー・ワークショップ
4. 発表年 2021年

1. 発表者名 川上 哲志
2. 発表標題 ナノフォトニックニューラルネットワークアクセラレータ
3. 学会等名 通信方式研究会 (招待講演)
4. 発表年 2021年

1. 発表者名 井上 弘士, 川上 哲志
2. 発表標題 光デバイス × アーキテクチャ：次世代光コンピューティングの開拓
3. 学会等名 フォトニックネットワーク研究会（招待講演）
4. 発表年 2021年

1. 発表者名 川上 哲志
2. 発表標題 集積ナノフォトニクスを活用した光電融合コンピューティングアーキテクチャ
3. 学会等名 System NanoTechnology研究会（招待講演）
4. 発表年 2021年

1. 発表者名 Koji Inoue, Satoshi Kawakami
2. 発表標題 Towards Ultra-Efficient Nanophotonic AI Acceleration ~ from a computer architecture perspective ~
3. 学会等名 Optical Fiber Communication Conference（招待講演）（国際学会）
4. 発表年 2021年

1. 発表者名 川上 哲志
2. 発表標題 コンピュータアーキテクチャの視点からみる光デバイス活用の課題と可能性
3. 学会等名 非ノイマン型情報処理へ向けたデバイス技術分科会（招待講演）
4. 発表年 2021年

1. 発表者名 川上哲志
2. 発表標題 計算機科学屋からみるフォトニックニューラルネットワークアクセラレータ
3. 学会等名 電子情報通信学会, フォトニックネットワーク研究会 (招待講演)
4. 発表年 2020年

1. 発表者名 川上哲志
2. 発表標題 ナノフォトニックニューラルネットワークアクセラレータ
3. 学会等名 電子情報通信学会, 通信方式研究会 (招待講演)
4. 発表年 2020年

1. 発表者名 井上弘士, 川上哲志
2. 発表標題 光デバイス × アーキテクチャ: 次世代光コンピューティングの開拓
3. 学会等名 電子情報通信学会, フォトニックネットワーク研究会 (招待講演)
4. 発表年 2020年

1. 発表者名 上野 麟, 谷本 輝夫, 後藤 孝行, 丸岡 晃, 川上 哲志, 小野 貴継, 飯塚 拓郎, 井上 弘士
2. 発表標題 オイラー動画画像誇張処理を対象とした CPU-FPGAハイブリッドシステムの実装と評価
3. 学会等名 情報処理学会, アーキテクチャ研究会
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------